REMARKS

Claims 1-13, 15-20, and 22-26 are all the claims pending in the application. Claims 14 and 21 have been cancelled herewith without prejudice or disclaimer. Claims 1-26 are objected to because of informalities. Moreover, claims 1-26 stand rejected on informalities. Claims 1, 3, 4, 6, 7, 13, 15, 16, 18-20, 22, 23, 25, and 26 stand rejected on prior art grounds. Claims 8-12 contain allowable subject matter. Claims 2, 5, 14, 17, 21, and 24 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form. Moreover, the specification and drawings are objected to. Applicants respectfully traverse the rejections based on the following discussion.

I. Information Disclosure Statement

The Office Action requests submission of an information disclosure statement containing each article referred to in the specification on pages 12 and 25. As such, Applicants are submitting an information disclosure statement concurrently with the filing of this amendment, including a copy of each of the references cited.

II. The Objections to the Drawings

The drawings are objected to under 37 C.F.R. §1.83(a) because the some of the features of the invention are not shown, such as the convex programming formulation and the objective function. Moreover, the drawings are further objected to because the Office

Action indicates it is not clear what feature Figures 2A, 2B, 2C, and 2D illustrate, and because the vertical axis in Figure 2C and 2D are not labeled.

Applicants herein submit newly added FIG. 5, which is a flowchart illustrating the preferred method of practicing the invention. The flowchart clearly refers to some of the features of the invention such as the convex programming formulation (shown in step 112) and the objective function (shown in step 114). Moreover, the specification has been amended to refer to FIG. 5 in the section entitled, "DESCRIPTION OF THE DRAWINGS", and the flowchart is further detailed in a new paragraph added in the specification (see amended specification). Furthermore, the original specification provides ample reference to the specifics summarized in the newly added flowchart throughout the specification, and in particular on page 11, line 4 through page 12, line 13 of the original specification. Thus, no new matter is added.

With regard to Figures 2A-2D, FIGs. 2C and 2D are amended herein to provide a label for the vertical axis. The vertical axis label in each figure is macro-p, which has been shown in red ink in the drawings included herein. As for the features shown in FIGs. 2A-2D, on page 23, line 3 through page 24, line19, the descriptions of the features shown in FIGs. 2A-2D are clearly given. Generally, the figures show graphical results of an example using the invention. Moreover, the specification has been amended in the section entitled, "DESCRIPTION OF THE DRAWINGS" to reflect this generalized explanation of FIGs. 2A-2D. Specifically, FIGs. 2A-2D indicate the position of the optimal weight tuples, wherein each data object x is represented as a m-tuple. This is

clearly discussed on pages 23-24 of the original specification, and those of ordinary skill in the art would readily understand the significance of the data plots illustrating an example of the invention as elaborated on in the specification. Applicants shall file new formal drawings, in the event they are necessary, for the amended drawings upon indication of allowance of the application. Therefore, the Examiner is respectfully requested to reconsider and withdraw this objection, to accept the proposed drawing changes, and to accept the newly added FIG.5.

III.   The Objection to the Specification

The specification as originally filed is objected to because of informalities. In accordance with the request in the Office Action, the Applicants have reviewed the entire specification and have made appropriate corrections in several areas, and as such, include a substitute specification herewith. Therefore, the Examiner is respectfully requested to substitute this specification for the specification originally filed with the application, and to accept its changes. The changes made are to provide for clarification, to correct typographical errors, and to correct grammatical errors contained in the original specification.

For example, the embedded hyperlinks are herein deleted as well as the duplicate paragraph on page 6 of the original specification. Moreover, all references to "Figures" have been effectively removed, and in its place, "FIG." has been inserted to correspond with the drawings.

The Office Action states that the original specification contains an incomplete description for "HEART (resp. ADULT) data" and "macro-p". However, the original specification clearly describes what these terms represent. For example, on pages 23 through 24, the original specification describes an example using the invention using HEART data, which is a data set consisting of 270 data instances. Every instance consists of 7 numerical and 6 categorical features. More specifically, the data set has two classes: the absence and presence of heart disease, where 55.56% of the data set consists of individuals who do not have heart disease (the absence of heart disease) and 44.44% of the data set consists of individuals who do have heart disease (the presence of heart disease). Moreover, the ADULT data set consists of 32,561 data instances, wherein every instance consists of 6 numerical and 8 categorical features. This data set has two classes: those with income less than or equal to $50,000, and those with income more than $50,000, where 75.22% of the data set consists of individuals (adults in the 1994 Census database) having an income less than or equal to $50,000 and 24.78% of the data set consists of individuals (adults in the 1994 Census database) having an income more than $50,000. The acronym "resp." as used throughout the text of the specification refers to "respectively". Thus, FIG. 2C illustrates the HEART data set and FIG. 2D illustrates the ADULT data set.

The phrase "macro-p" is also clearly defined on page 22 of the original specification. Performance averages across classes are calculated using macro-precision (macro-p), macro-recall (macro-r), micro-precision (micro-p), and micro-recall (micro-r).

Mathematically, macro-p is given as:

$$\frac{1}{c}\sum_{t=1}^{c} p_t$$

Those skilled in the art would readily understand the mathematical expression provided above as well as its significance in data clustering and optimization, especially when read in the context of the entire specification.

Next, the Office Action posits the difference between Q1xQ2 and Q1xQ2xQ3. Again, the original specification clearly defined what the above represents. For example, Q1xQ2 represents the objective function. For the HEART and ADULT data, FIG. 2A and 2B, respectively show a plot of the objective function $Q_1 \times Q_2$ in equation (6) versus the weight α_1. For the HEART and ADULT data sets, the objective function is minimized by the weights (0.12, 0.88) and (0.11, 0.89), respectively.

For the HEART and ADULT data, FIG. 2C and 2D, respectively show a plot of macro-p (resp. micro-p, macro-p, and macro-r) versus the weight α_1. By comparing FIG. 2A with FIG. 2C and FIG.2B with FIG. 2D, it can be seen that, roughly, macro-p (resp. micro-p, macro-p, and macro-r) are negatively correlated with the objective function $Q_1 \times Q_2$ and that, in fact, the optimal weight tuples achieve nearly optimal precision and recall. In conclusion, optimizing the objective function $Q_1 \times Q_2$ leads, reassuringly, to optimizing the precision/recall performance, thus leads to good clusterings and a final solution.

Q1xQ2xQ3 represents the objective function in a three-vector example, as further discussed on pages 25-28 of the original specification. In contrast Q1xQ2 represents the objective function in a two-vector example. Therefore, the difference between the two objective functions, respectively, are dimensional differences. Moreover, the objective function is defined throughout the specification as essentially the means by which the feature weights of the feature weights are optimized in order to produce the final clustering solution that simultaneously minimizes average intra-cluster dispersion and maximizes average inter-cluster dispersion along all of the heterogeneous feature spaces. Moreover, the objective function is further defined mathematically in the specification (see for example equation (5) on page 20 of the original specification).

Similarly, the convex programming formulation is defined mathematically in equations (1) and (2) provided on page 15 of the original specification. Moreover, the Office Action readily admits that convex programming formulations are known in the art, and uses U.S. Patent 5,596,719 issued to Ramakrishnan et al. as a basis for this argument. Moreover, those skilled in the art of data clustering and optimization would understand the significance of the convex programming formulation provided by the claimed invention. In view of the foregoing, the Examiner is respectfully requested to reconsider and withdraw the objections to the specification.

IV. **The Objections to the Claims**

Claims 1-26 are objected to because of informalities. As such, the Applicants

have amended the claims in accordance with the suggestions made by the Examiner to remove the offending informalities. Therefore, the Examiner is respectfully requested to reconsider and withdraw the objections to the claims.

V.   The Claim Rejections

A.   The 35 U.S.C. § 112, first paragraph Rejections

Claims 1-26 are rejected under 35 U.S.C. § 112, first paragraph, as containing subject matter which was not described in the specification in such a way as to enable one skilled in the art to which it pertains, or with which it is most nearly connected, to make and/or use the invention. Specifically, the Office Action states that claims 1-26 recite a "convex programming" and an "objective function" as the main components for the invention, but that the specification does not contain any definition of those terms. Applicants respectfully traverse these rejections and strongly refute the assertion that the specification does not adequately provide the definition of the terms "convex programming" and "objective function". First, as amended both the specification and claims are now written with the terms convex programming and objective function not indicated in double quotation marks. Second, as indicated above, the specification does in fact refer to each term and provides an ample discussion of what the term represents and the context in which each term is used in the claimed invention.

To reiterate, Q1xQ2xQ3 represents the objective function in a three-vector example, as further discussed on pages 25-28 of the original specification. In contrast

Q1xQ2 represents the objective function in a two-vector example. Therefore, the difference between the two objective functions, respectively, are dimensional differences. Moreover, the objective function is defined throughout the specification as essentially the means by which the feature weights of the feature weights are optimized in order to produce the final clustering solution that simultaneously minimizes average intra-cluster dispersion and maximizes average inter-cluster dispersion along all of the heterogeneous feature spaces. Moreover, the objective function is further defined mathematically in the specification (see for example equation (5) on page 20 of the original specification).

Similarly, the convex programming formulation is defined mathematically in equations (1) and (2) provided on page 15 of the original specification. Moreover, the Office Action readily admits that convex programming formulations are known in the art, and uses U.S. Patent 5,596,719 issued to Ramakrishnan et al. as a basis for this argument. Moreover, those skilled in the art of data clustering and optimization would understand the significance of the convex programming formulation provided by the claimed invention. In view of the foregoing, the Examiner is respectfully requested to reconsider and withdraw these rejections.

B.   The 35 U.S.C. § 112, second paragraph Rejections

Claims 1-26 are rejected under 35 U.S.C. § 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which the applicant regards as the invention because of informalities. Specifically, the Office

Action indicates that claims 1, 2, 5, 8, 10, 13, 14, 17, 20, 21, and 24 contain terms in double quotation marks not explicitly defined in the specification. Moreover, claims 5 and 17 are rejected for containing terms lacking proper antecedent basis. As such, claims 1, 2, 5, 8, 10, 13, 17, 20, and 24 are amended to remove the double quotation marks. Also, claims 14 and 21 are cancelled. Additionally and for the reasons described above, the Applicants traverse the assertion that certain terms recited in the claims are not adequately defined in the specification. As fully described above all of the terms recited in the claims have been fully described, discussed, defined (mathematically or otherwise) in the specification. In the interest of brevity, the Applicants' arguments demonstrating this will not be reproduced here. However, the Applicants refer to the arguments given to the objection to the specification and to the claim rejections for proof of such proper discussion and description of the claimed language. Furthermore, claims 5 and 17 are amended herein to provided proper antecedent basis for the offending language.

C.   The Prior Art Rejections under 35 U.S.C. § 103(a)

Claims 1, 3, 5, 6, 7, 13, 15, 16, 18-20, 22, 23, 25, and 26 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Fayyad et al. (United States Patent No. 6,115,719), hereinafter referred to as "Fayyad" in view of Ramakrishnan et al. (United States Patent No. 5,596,719), hereinafter referred to as "Ramakrishnan". Specifically, the Office Action indicates that Fayyad discloses some of the elements of claimed invention, but not all. The Office Action suggests that Fayyad does not specifically show that a

convex programming formulation is used. However, the Office Action suggests that Ramakrishnan uses convex programming to find optimum solutions, and as such it would be obvious to one of ordinary skill in the art to include the claimed convex programming formulation and feature weights selection while implementing the method of Fayyad in order to take advantage of a well known optimization technique. Additionally, the Office Action states that Fayyad does not specifically show analyzing word data and feature vectors comprising multiple-word frequencies of the data records, but that it is well known in the art to cluster documents using a word frequency. Therefore, the Office Action concludes that it would have been obvious to one of ordinary skill in the art to include the claimed feature while implementing the method of Fayyad depending on the user's requirement. The Applicants traverse these rejections based on the following discussion.

1. The Prior Art References

a. The Fayyad Reference

Fayyad teaches a method that takes an initial condition and efficiently produces a refined starting condition. The method is applied to the K-means clustering algorithm and shows that refined initial starting points indeed lead to improved solutions. The technique can be used as an initializer for other clustering solutions and is based on an efficient technique for estimating the modes of a distribution and runs in time guaranteed to be less than overall clustering time for large data sets. The method is also scalable and

hence can be efficiently used on huge databases to refine starting points for scalable clustering algorithms in data mining applications.

b.   The Ramakrishnan Reference

Ramakrishnan teaches a method and apparatus for assigning link "distance" metrics that result in near optimal routing for a network formed of nodes (routers) and links, where each link has a capacity associated with it, and where source-destination flows are given. The routing optimality is measured with respect to some objection function (e.g., average network delay).

2.   Applicants Response

As amended, the claimed invention is patentable over the supposed combination of Fayyad with Ramakrishnan. Specifically, the prior art of record does not disclose or make obvious "a method for evaluating and outputting a final clustering solution for a plurality of multi-dimensional data records, said data records having multiple, heterogeneous feature spaces represented by feature vectors, said method comprising: defining a distortion between two feature vectors as a weighted sum of distortion measures on components of said feature vectors; clustering said multi-dimensional data records into k-clusters using a convex programming formulation; selecting feature weights of said feature vectors, and <u>minimizing distortion of said k-clusters</u>." In fact, neither Fayyad nor Ramakrishnan disclose minimizing distortion of k-clusters at all. Nor

would it be obvious to combine such a feature with the other features of the claimed invention because both Fayyad and Ramakrishnan of functionally complete techniques in their own right, and to combine them and then add to them the additional feature of minimizing distortion of the k-clusters would presuppose an unobvious combination and an unmotivated tendency for someone of ordinary skill in the art.

Therefore, the claimed invention is patentably distinct from either Fayyad or Ramakrishnan, whether taken alone or in combination with one another, and moreover, the invention is unobvious in light of the teachings of both Fayyad and Ramakrishnan. Thus, claims 1, 3, 4, 6, 7, 13, 15, 16, 18-20, 22, 23, 25, and 26 are patentably distinct over Fayyad in combination with Ramakrishnan and are in condition for allowance. In view of the foregoing, the Examiner is respectfully requested to reconsider and withdraw this rejection.

The Office Action indicates that claims 8-12 would be allowable if rewritten to overcome the rejections under 35 U.S.C. § 112, first and second paragraphs and to overcome the objection to the claims as well. As such, claims 8-12 have been amended in the manner suggested by the Examiner, and it is respectfully requested that these amended claims be placed in condition for immediate allowance.

Moreover, the Office Action indicates that claims 2, 5, 14, 17, 21, and 24 would be allowable if rewritten in independent form to include all of the limitations of the base claim and any intervening claims and to overcome the rejection under 35 U.S.C. § 112, first and second paragraphs and to overcome the objection to the claims as well. As such,

claim 1 has been amended to further differentiate it from the prior art of record. Thus, claims 2 and 5, in their dependent state, are also further differentiated from the prior art. Furthermore, claims 13 and 20 have been amended to include the limitations of claims 14 and 21, respectively, which are hereby cancelled. Therefore, amended claims 13 and 20 are in condition for immediate allowance. As such, because amended claims 13 and 20 contain subject matter which the Examiner has deemed allowable, it follows that the dependent claims which depend thereon, respectively, also contain allowable subject matter. Therefore, dependent claims 15-19, and 22-26 are in condition for immediate allowance.

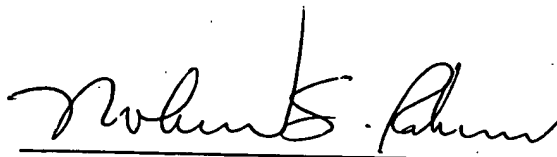## VI. Formal Matters and Conclusion

Therefore, Applicants respectfully submit that amended independent claims 1, 8, 13, and 20 are patentable over Fayyad, even if combined with Ramakrishnan. Furthermore, dependent claims 2-7, 9-12, 15-19, and 22-26 are similarly patentable, not only by virtue of their dependency from a patentable independent claim, but also by virtue of the additional features of the invention they define. In view of the foregoing, Applicants submit that claims 1-13, 15-20, and 22-26, all the claims presently pending in the application, are patentably distinct from the prior art of record and are in condition for allowance. Furthermore, no new matter has been added. The Examiner is respectfully requested to pass the above application to issue at the earliest possible time.

Should the Examiner find the application to be other than in condition for

allowance, the Examiner is requested to contact the undersigned at the local telephone number listed below to discuss any other changes deemed necessary. Please charge any deficiencies and credit any overpayments to Attorney's Deposit Account Number 09-0456.

Respectfully submitted,

Dated: 5-13-03

Mohammad S. Rahman
Reg. No. 43,029

McGinn & Gibb, P.L.L.C.
2568-A Riva Road
Suite 304
Annapolis, MD 21401
(301) 261-8625
Customer Number: 28211